

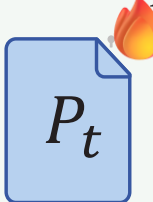
Dataset



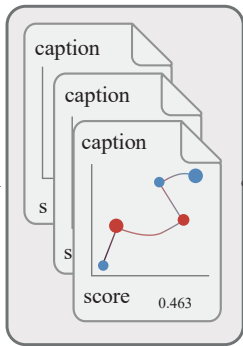
Model

Frozen
Qwen3-VL 32B 

Trainable Prompt

Markdown File 
 P_t

1000 Training Data
Rollout batch



Rubric Scorer

GPT-5 provides the scores and analysis of each sample in terms of the following dimensions:

- Completeness
- Redundancy
- Coherence
- OCR fidelity

Scores and
Analysis

Prompt Modifier

Github copilot acts as the prompt modifier:

- Modify the prompt with minimal edit
- Copilot can modify the code with human approve

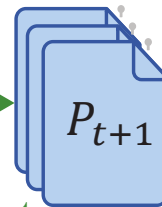


Full tool set of
coding agents

200 Validation Data
Validation batch

Validation
gate

Current
prompt



Accept

Reject



Logs of Edit History

Best
prompt